

TMH:jlb 8/11/03 212072.doc
PATENT

Attorney Reference Number 4810-58741
Application Number 09/806,708

AMENDMENTS TO THE SPECIFICATION

Please replace the paragraph beginning at page 7, line 1, with the following amended paragraph:

--Optimal alignment of sequences for comparisons of similarity may be automated using a variety of algorithms, such as the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math* 2: 482, the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, and the computerized implementations of these algorithms (such as GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, WI, U.S.A.). Sequence similarity may also be determined using the BLAST algorithm, described in Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10 (using the published default settings). Software and instructions for performing BLAST analysis may be available through the National Center for Biotechnology Information in the United States (including the programs BLASTP, BLASTN, BLASTX, TBLASTN and TBLASTX that may be available through the internet at the NCBI website. ~~www.ncbi.nlm.nih.gov/~~ The BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database (reference) sequence. T is referred to as the neighborhood word score threshold. Initial neighborhood word hits act as seeds for initiating searches to find longer HSPs. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction is halted when the following parameters are met: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program may use as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919), a gap existence cost of 11, a per residue gap cost of 1, a lambda ratio of 0.85, alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands. One measure of the statistical similarity between two sequences using the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match

TMH:jlb 8/11/03 212072.doc
PATENT

Attorney Reference Number 4810-58741
Application Number 09/806,708

between two nucleotide or amino acid sequences would occur by chance. In alternative embodiments of the invention, nucleotide or amino acid sequences are considered substantially identical if the smallest sum probability in a comparison of the test sequences is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. In the PSI-BLAST implementation of the BLAST algorithm, an expect value for inclusion in PSI-BLAST iteration may be 0.001 (Altschul *et al.* (1997), *Nucleic Acids Res.* 25:3389-3402). Searching parameters may be varied to obtain potentially homologous sequences from database searches. --